Final Report | Capstone Project – The Battle of Neighbourhoods | Choosing the best bakery location in Toronto

1. Introduction:

Toronto is said to be the largest city of Canada. Many immigrant cultures have come to Toronto and have brought their traditions, languages, food, and music to the city. This shows that Toronto is the most ethnically diverse city. Besides, the diverse culture that is brought into the city, Toronto is also a business-minded, conscientious, socially progressive, and pluralistic city. It is a city of many museums, theatres, festival events and sports activities.

Thus, the purpose of this project is to provide enough information for someone who would like to open up a bakery in the city of Toronto. The idea is to help someone make a smart and efficient decision on selecting a great location for starting a bakery in Toronto. Fast forward in time, this will help the owner of the bakery to feel he/she made the right decision and hopefully have the bakery bringing in more customers year on year, and simultaneously growing the business over time.

So, as part of this project, we will list and visualise all major parts of Toronto City that has great bakeries spread across the city.

2. Data Section

Data Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

We will use the dataset which we scrapped from Wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes.

Foursquare API Data:

We will need data about venues that are related to bakeries in different neighbourhoods. In order to gain that information, we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighbourhoods, we then connect to the Foursquare API to gather information about venues inside each neighbourhood. For each neighbourhood, we have chosen the radius to be 700 meters.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighbourhood
2. Neighbourhood Latitude
3. Neighbourhood Longitude
4. Venue

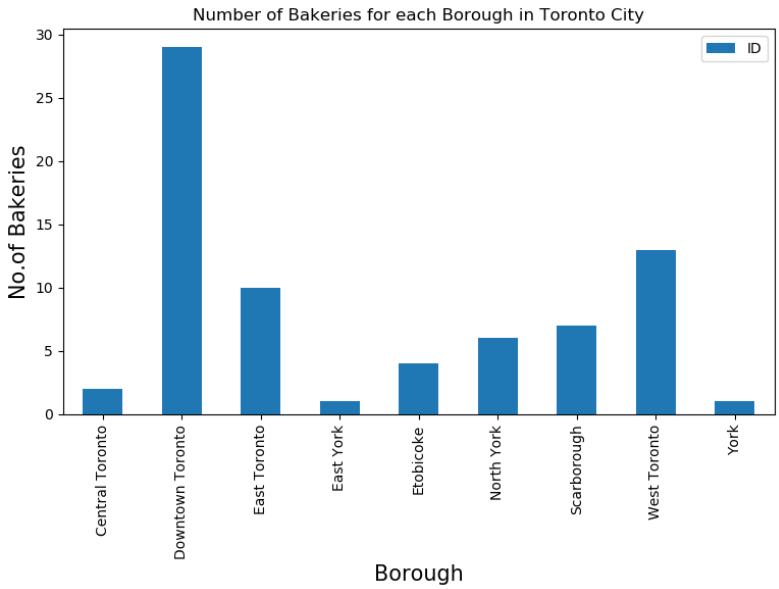
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

3. Methodology Section

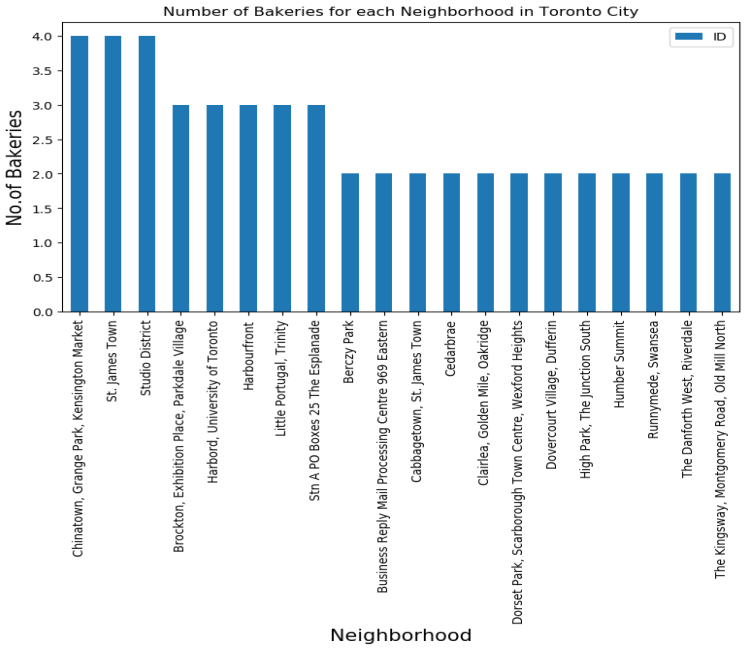
1. Collect the Toronto city data from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M.
2. Using Foursquare API, we will find all venues for each neighbourhood. Also using credentials of Foursquare API features of near-by places of the neighbourhoods would be mined. Due to http request limitations the number of places per neighbourhood parameter would reasonably be set to 100 and the radius parameter would be set to 700.
3. Filter out all venues that are bakeries.
4. Find ratings, and the number of bakeries found in each neighbourhood using Foursquare API.
5. We use the rating to Using rating for each restaurant, we will sort that data.
6. Visualize the Ranking of neighbourhoods using folium library(python). This can determine where will there be competition and which all neighbourhoods will have potential to start a bakery.

4. Results Section

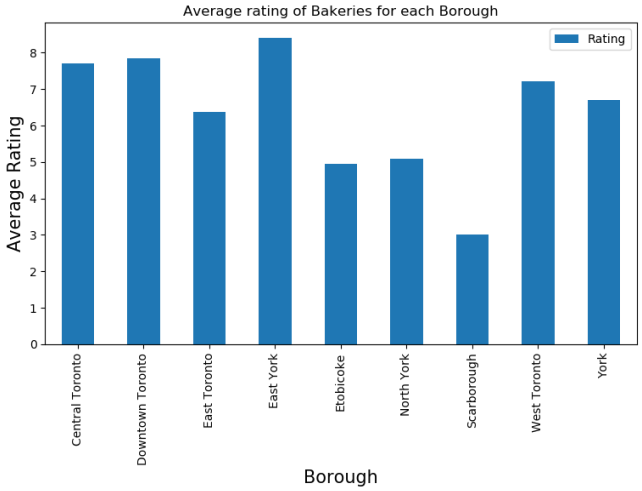
Number of Bakeries for each Borough in Toronto City:



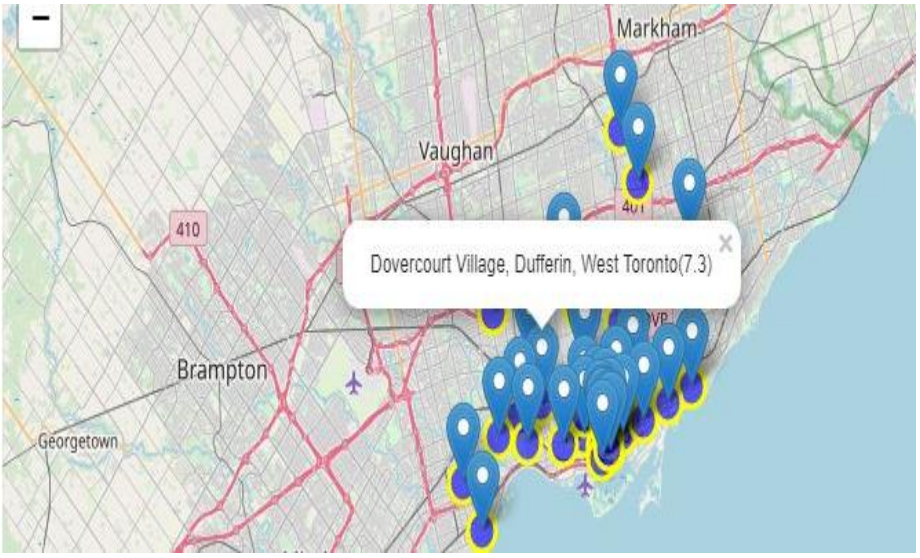
Number of Bakeries for each Neighbourhood in Toronto City:



Average rating of Bakeries for each Borough:



Map to visualize the Neighbourhood and borough with the average rating of the bakery:



Foursquare API:

This project has used Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

5. Discussion Section

The purpose of this project was to provide enough information for someone who would like to open a bakery in the city of Toronto. The idea was to help someone make a smart and efficient decision on selecting a great location for starting a bakery in Toronto.

It showed that Downtown Toronto had the greatest number of bakeries in Toronto, which was close to 30 bakeries. East York showed highest average rating, in terms of great bakeries to go to. As a recommendation it is best to start a bakery shop at East York, because it has the least number of bakeries in the Borough and it has one of the best ratings among the Boroughs.

6. Conclusion Section

In this project, we used charts to visualize the average ratings that can be used to determine where a bakery can be located when starting a business.

I feel rewarded with the efforts and believe this course with all the topics covered is well worthy of appreciation. This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools. The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.

Libraries Which are Used to Develop the Project:

Pandas: For creating and manipulating dataframes.

Folium: Python visualization library would be used to visualize the neighbourhoods cluster distribution of using interactive leaflet map.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Data.

Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.